Moore, J.C., and Klaus, A. (Eds.)
Proceedings of the Ocean Drilling Program, Scientific Results Volume 171A

2. Data Report: Logging While Drilling Data Analysis of Leg 171A, a Multivariate Statistical Approach[1]

C. Bücker,[2] J. Shimeld,[3] S. Hunze,[2] and W. Brückmann[4]

ABSTRACT

In the northern Barbados accretionary wedge, several Deep Sea Drilling Project (DSDP) and Ocean Drilling Program (ODP) legs (DSDP Leg 78 and ODP Legs 110, 156, and 171A) targeted the décollement and the seaward extension of the décollement, the proto-décollement. During Leg 171A, the logging while drilling (LWD) technique was used to determine the physical properties variations along a profile across the deformation front. Because of the unstable borehole conditions in accretionary wedges, LWD is the most effective method for the measurements of physical properties in these poorly consolidated sediments. LWD data are acquired just above the drill bit a few minutes after the formation has been drilled, yielding measurements as close to in situ conditions as possible.

The large amount of LWD data and the demand for a quick, objective, and reliable evaluation calls for the application of multivariate statistical methods. The multivariate factor analysis is a method of reducing the amount of logging data while giving them a new integrated meaning with no loss of important information, resulting in factor logs that are helpful tools for further interpretation. The cluster analysis of the two or three most significant factors proved to be a useful and objective method to identify and confirm significant logging units. The main objective of the application of multivariate statistical methods in this study is twofold. First, Leg 171A was a stand-alone logging leg, where no cores were retrieved. The factor analysis was used as an objective tool for a classification of the drilled sequences based on their physical and chemical properties. The new factor logs mirror the

[1]Bücker, C., Shimeld, J., Hunze, S., and Brückmann, W., 2000. Data report: LWD data analysis of Leg 171A, a multivariate statistical approach. *In* Moore, J.C., and Klaus, A. (Eds.), *Proc. ODP, Sci. Results,* 171A, 1–29 [Online]. Available from World Wide Web: <http://www-odp.tamu.edu/publications/171A_SR/VOLUME/CHAPTERS/SR171A02.PDF>. [Cited YYYY-MM-DD]
[2]Joint Geoscientific Research Institute, 30631 Hannover, Stilleweg 2, Federal Republic of Germany. Correspondence author: **c.buecker@gga-hannover.de**
[3]Geological Survey of Canada (Atlantic), PO Box 1006, Dartmouth, NS B2Y 4A2, Canada
[4]GEOMAR Research Center for Marine Geosciences, University of Kiel, Wischhofstrasse 1-3, 24148 Kiel, Federal Republic of Germany.

Initial receipt: 11 May 1999
Acceptance: 22 November 1999
Web publication: 2 August 2000
Ms 171ASR-103

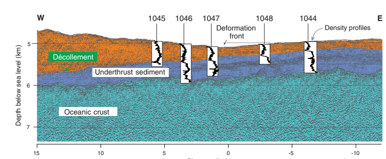
basic processes behind the measured geophysical properties and make them easier to interpret. Second, in the succeeding cluster analysis, similar geophysical properties are grouped into one cluster, reflecting one logging unit. These objectively defined logging units can be compared to statistical electrofacies, which are helpful in differentiating lithologic characterizations. In particular for LWD measurements, the multivariate statistical methods of factor and cluster analysis are helpful tools for a fast, reliable, and objective definition of logging units, which should be considered for future legs.

INTRODUCTION

Barbados Accretionary Wedge

The northern Barbados accretionary wedge is located along a convergent margin that is actively accreting oceanic sediments. It develops where Upper Cretaceous Atlantic Ocean crust underthrusts the Caribbean plate in a western direction. The accretionary wedge consists of Quaternary to Miocene calcareous mud, mudstone, and claystone (Moore et al., 1998). Detailed knowledge of the Barbados accretionary wedge has been obtained through several Deep Sea Drilling Project (DSDP) and Ocean Drilling Project (ODP) legs (DSDP Leg 78 and ODP Legs 110, 156, and 171A). At the Leg 171A sites, the detachment between the underthrusting oceanic plate and the accretionary prism, which is known as the décollement zone, occurs at ~200–400 m below seafloor (mbsf) (Fig. F1) (for a borehole localization map see Moore, Klaus, et al., 1998). The underthrust sequence consists of lower Miocene to Oligocene mudstone, claystone, and turbidites down to 820 mbsf (Moore, Klaus, et al., 1998).

F1. Seismic depth section from west of Site 1045 to east of Site 1044, p. 12.



Deformation and fluid flow in this accretionary prism change the physical properties of the sediments. In some cases, changes in physical properties are localized along discrete faults in response to overpressuring and fluid migration, whereas, in other cases, changes in physical properties reflect variations in the broader stress regime (Shipley et al., 1994). The evolution of these physical properties cannot be comprehensively derived from recovered cores because of elastic rebound and microcracking effects.

One of the main objectives during Leg 171A was to map and understand the evolution of changes in physical properties within the accretionary wedge (Moore, Klaus, et al., 1998). Logging with conventional open-hole wireline logs proved difficult to impossible during previous legs (Legs 110 and 156) because boreholes penetrating the unconsolidated sediments were too unstable, especially near the décollement zone (Jurado et al., 1997). The logging while drilling (LWD) technique was used for the first time by the Ocean Drilling Program during Leg 156. During Leg 171A, this technology was used solely for borehole measurements.

Logging While Drilling Technique

Sensors in the LWD tool are located inside the drill string, 3–13 m above the drill bit. This allows geophysical measurements of the formation to be made shortly after the drill bit has penetrated it and before the borehole is affected by continued drilling or coring operations. Thus, the measurements are not influenced by borehole breakouts or washouts. In addition, because measurement occurs within minutes of the hole being drilled, the effects of borehole wall infiltration are minimized. Geophysical analysis with LWD tools may become a routine procedure in soft, unstable, or overpressured sediments. In a single logging run, data for up to 10 or more physical, chemical, and technical parameters can be obtained (Shipboard Scientific Party, 1998). The interpretation of the resulting data matrices requires a profound geophysical and sedimentological background and can benefit from sophisticated statistical operations.

Multivariate Analyses

Multivariate statistical analyses of LWD data have not been common. But the large amount of data from LWD measurements and the demand for a fast, reliable, and objective evaluation and interpretation makes the application of multivariate statistical methods ideal. In this study, the multivariate statistics procedures of factor and cluster analysis are used to obtain quick results from the LWD measurements. Factor analysis is used to rescale and reduce the original data set and to derive a deeper insight into the background processes. Cluster analysis is used to define electrofacies in as objective a manner as possible, which is particularly important for Leg 171A because no cores were collected.

LWD DATA AND QUALITY

A complete set of LWD data were recorded in all Leg 171A holes using the Schlumberger-Anadrill compensated dual resistivity (CDR) and compensated density neutron (CDN) tools. Although these tools differ slightly from conventional wireline logging tools, they are based on the same physical principles and results comparable to wireline logging can be obtained. One of the main differences is that the data are not recorded with depth but with time. The downhole data acquisition systems are synchronized with a system on the rig that monitors time and drilling depth. After completion of the drilling, the data are downloaded from memory chips in the tools and the time-depth conversion is made. In contrast to conventional wireline logging data, depth mismatches between different logging runs are impossible because the data are all obtained during a single logging run.

A full description of the principles and measurements performed by the LWD tools is given by Anadrill-Schlumberger (1993) and Shipboard Scientific Party (1998). All Leg 171A holes (1044A, 1045A, 1046A, 1047A, and 1048A) were successfully logged with both the CDR and the CDN tools, and the data are considered to be of overall good quality. This is the most complete and comprehensive data set of in situ geophysical measurements in an accretionary wedge drilled by ODP. Physical and chemical properties measured by the CDR and CDN tools include spectral gamma ray (GR); thorium, uranium, and potassium content (Th, U, and K); computed gamma ray (CGR); formation bulk density (ROMT); photoelectric effect (PEF); differential caliper; attenuation resistivity (ATR); phase shift resistivity (PSR); and neutron porosity (TNPH). Additional parameters of geotechnical significance, such as the rate of penetration and weight on bit, are also collected. The radius of investigation and vertical resolution of LWD logging tools vary depending on the measuring principle and measured property. For example, the PSR curve provides shallow resistivity estimates in comparison to the deeper reading ATR curve. The PSR and ATR measurements are most accurate within low-resistivity formations (<2 m) (Anadrill-Schlumberger, 1993), which is the typical case in accretionary wedges where sediments are unconsolidated and porosities tend to be high (>40%). The TNPH measurement responds not simply to formation porosity, but also to the hydrogen content within the bulk rock. Thus, in clay-rich formations TNPH records the combined effect of porosity and clay content. Chemical elements with large neutron cross sections like gadolinium may have also an effect on the neutron porosity readings. Unfortunately, no gadolinium content measurements were available until now for the Barbados accretionary wedge sediments. TNPH measurements are most accurate in formations with porosities not >40% (Theys, 1991). Porosities in the Barbados accretionary wedge are as high as 70%, resulting in noisy and scattered TNPH data.

Statistical Methods and Theoretical Background

A description of the basic onboard data treatment is given in the *Initial Reports* volume of Leg 171A (Moore, Klaus, et al., 1998). In this volume, a detailed and expanded procedure of data processing is described and documented. Excellent reviews of general statistical techniques, their use in geosciences, and examples in borehole geophysics are given by Backhaus et al. (1996), Brown (1998), Bucheb and Evans (1994), Davis (1986), Doveton (1994), Elek (1990), Harvey and Lovell (1989), Harvey et al. (1990), Howarth and Sinding-Larsen (1983), and Rider (1996).

Data Preparation

The statistical methods described in this paper require that the observational data set (i.e., the geophysical measurements) be normally distributed. When this is not the case, the observations should be transformed so that they more closely follow a normal distribution. For example, electrical resistivities often appear to follow a lognormal distribution, and application of a logarithmic transform will yield observations that are more normally distributed. Erroneous values, when they can be clearly identified, must also be omitted from the analysis. Fortunately, LWD generally provides large, reliable data sets so that this editing procedure has little negative effect on the analysis.

Finally, before beginning the statistical analysis, the observational data should be rescaled by subtracting the mean and dividing by the standard deviation (i.e., a "standardization" of data). The resulting values will be dimensionless and will have a mean of zero and a standard deviation of 1. This permits comparison between all the observations regardless of their original scaling.

Factor Analysis

Factor analysis (FA) is a technique for examining the interrelationships among a set of observations. It is used to derive a subset of uncorrelated variables called factors that adequately explain the variance observed in the original observational data set (Brown, 1998). Often such analysis reveals structure in the data set by identifying which observations are most strongly correlated. Interpretation of these correlations contributes to understanding of the underlying processes that are being measured. A significant advantage of FA is that the number of variables can be dramatically reduced without losing important information. In other words, the dimensionality of the observational data set can be reduced. Half a dozen or more interrelated variables might be reduced to perhaps two or three factors that account for nearly all the variance in the original data set. Visualization of two or three factors is much simpler than visualization of the entire data set.

When comparing German and U.S. literature, FA is sometimes confused with the principal component analysis (PCA). But there is a significant difference between the two techniques. Strictly speaking, principle components are the eigenvectors of the covariance or correlation matrix of the observations. Statistical considerations such as probability or hypothesis testing are not included in PCA (Davis, 1986). Often though, PCA forms the starting point for FA. In FA, a series of assumptions are made regarding the nature of the parent population from which the samples (i.e., observations) are derived. For example, the observations are assumed to follow a normal distribution. Such assumptions provide the rationale for the operations that are performed and the manner in which the results are interpreted (Davis, 1986).

Another way of explaining the difference between FA and PCA lies in the variance of variables (communality) that is analyzed. Under FA, attempts are made to estimate and eliminate variance caused by error and variance that is unique to each variable (Brown, 1998). The result of FA concentrates on variables with high communality values (Tabachnick and Fidell, 1989); only the variance that each variable shares with other observed variables is available for analysis and interpretation. In this investigation, the FA method is used because error and unique variances only confuse the picture of underlying processes and structures. Factors and factor loadings were calculated from the rescaled logging curves using standard R-mode factor analyses procedures (Davis, 1986) on the variables at each site. A Kaiser Varimax factor rotation (Davis, 1986) is applied because the matrix of factor loadings is often not unique or easily explained. The technique of factor calculation is that of extraction of the eigenvalues and the eigenvectors from the matrix of correlations, or covariances. With appropriate assumptions, the factor model is simply a linear combination of underlying variables and properties. A factor is taken as being significant for an underlying property if it adds a significant amount of variance, or in practical terms, if its eigenvalue is >1. Factors with eigenvalues <1 account for less variation than one of the initial variables.

Theoretically, because they are maximally uncorrelated, each factor represents an underlying rock property such as porosity, lithology, fracture density, water content, or clay type. This is not strictly the case, in reality, because there is obviously no precondition that the rock properties will themselves be uncorrelated. Indeed, it is possible to envision highly nonlinear interrelations between various rock properties like porosity, lithology, fracture density, fluid content, and clay type. As a first order interpretation though, FA provides an objective, rapid, and methodical approach for identifying major features of an observational data set. Also, since many borehole geophysical tools respond primarily to porosity and lithology, Elek (1990) argued that the first two factors (i.e., the two factors accounting for the highest degree of variance in the observations) derived from FA will also relate directly to porosity and lithology. This is a reasonable assertion when the interaction between various rock properties is known to be relatively simple. Such is the case at the Barbados accretionary wedge, where the sediments are unlithified, there is little secondary mineralization, the large-scale porosity trends are predictable, and the fluid content is well known. In many respects, accretionary wedge sediments are somewhat unique when compared, for example, to the typical variation in rock parameters that is encountered in petroleum industry applications.

Generally, for the Leg 171A LWD data sets, far more than 80% of the variance observed in the input variables can be described by the first two or three factors (Tables T1, T2, T3, T4, T5, T6, T7, T8, T9, T10). This means that the amount of explained variance is >80%, although the number of variables has been reduced from as much as 7 to 2 or 3.

Cluster Analysis

After performing FA, statistical electrofacies are defined using cluster analysis. Clustering techniques are generally used for grouping individuals or samples into a priori unknown groups. The objective of the cluster analysis is to separate the groups based on measured characteristics with the aim of maximizing the distance between groups. Hierarchical clustering methods yield a series of successive agglomerations of data points on the basis of successively coarser partitions. One of the most common methods of complete linkage hierarchical clustering is the Ward method (Davis 1986), which is used in this study.

Before applying the cluster analysis, the factor logs that are used as input variables are reduced to a 1-m depth interval using a finite-impulse response, low-pass antialiasing filter to reduce the number of data points. This step, although unnecessary, has two advantages. First, the cluster analysis, in particular when using the complete linkage hierarchical Ward method, is a very time and computer memory–consuming calculation procedure. Reducing the number of data points results in faster calculations. Second, this step was performed to get a cluster-log that does not show too many details (i.e., showing a new cluster every few centimeters). At the resolution shown in the figures, no loss of information is visible, justifying this reduction process. After this data reduction procedure on the factor logs, a complete linkage hierarchical cluster analysis using a Euclidean norm ("Ward method"; see Davis, 1986) was performed on the two or three decimated factors that accounted for the greatest amount of variance in the initial data set. This allowed the identification of statistical electrofacies, or logging units, with distinct combinations of rock physical and chemical properties (e.g., Serra, 1986). A dendrogram, which is a tree diagram showing similarity or connectivity between samples and clusters (e.g. Doveton, 1994) is used to decide how many clusters are significant and useful. For all sites, the number of clusters varies between 4 and 6. Of course, the likelihood for a greater number of significant clusters in deeper boreholes increases as the number of observations increases.

There are several commercial software packages that can be used to perform all the multivariate statistical methods described above. For this investigation we used WINSTAT 3.1 (Kalmia Software) and MVSP 3.0 (Kovach, 1998) on a PC platform under Windows NT 4.0 and 128 MB of RAM.

T1. Output and results of factor analysis for Site 1044, p. 20.

T2. Varimax factor loadings and communalities for Site 1044, p. 21.

T3. Output and results of factor analysis for Site 1045, p. 22.

T4. Varimax factor loadings and communalities for Site 1045, p. 23.

T5. Output and results of factor analysis for Site 1046, p. 24.

T6. Varimax factor loadings and communalities for Site 1046, p. 25.

T7. Output and results of factor analysis for Site 1047, p. 26.

T8. Varimax factor loadings and communalities for Site 1047, p. 27.

T9. Output and results of factor analysis for Site 1048, p. 28.

T10. Varimax factor loadings and communalities for Site 1048, p. 29.

APPLICATION OF MULTIVARIATE STATISTICAL METHODS

Factor Logs at Each Site

For the FA, all LWD data at each site except PSR and U were taken into account. The shallow resistivity data PSR were not used because they correlate strongly with the deep resistivity ATR and are nearly identical to ATR. This strong correlation would weigh resistivity too heavily compared to the remaining data. This identical behavior of PSR and ATR is an effect of measuring the formation only minutes after it has been drilled. There is no time that the mud can infiltrate into the formation. The deep resistivity was used rather than the shallow resistivity, because it is more likely to be representative of the undisturbed sediment away from the borehole. The U data were not used because they showed very low values over the entire borehole with little character and contained unreliable negative values (obviously due to the processing from the gamma-ray spectra). Accordingly, the computed gamma-ray CGR (GR with U portion subtracted) was also not used. Although the TNPH data showed a noisy and scattered character because of overall high porosities in the formation, a low-pass filtering of the data made them useful for the statistical analyses.

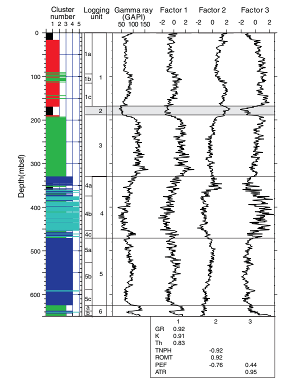
The results of the factor analysis of the LWD data with factor eigenvalues and factor loadings are given in Tables T1, T2, T3, T4, T5, T6, T7, T8, T9, and T10 and are graphically presented as factor logs in Figures F2, F3, F4, F5, and F6. Factor loadings >0.4 are significant and shown in bold in Tables T2, T4, T6, T8, and T10 and also at the bottom of Figures F2, F3, F4, F5, and F6. Except for Site 1045, where two factors could be extracted, three factors were extracted from the original data sets. At Site 1045, only ~40 m was logged below the décollement (Fig. F1). This reduced data set may have caused the smaller number of factors in this case.

The results at all sites show that factor 3 is closely related to the deep resistivity ATR (factor loadings >0.75). Accordingly at Site 1045 (Fig. F3), the ATR log is shown together with the two factor logs. The gamma-ray log is shown for all sites for comparison reasons.

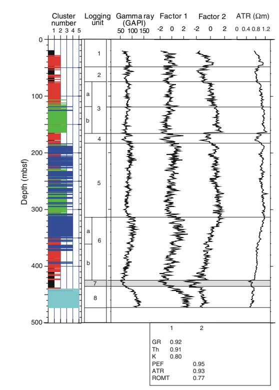
The factor analysis shows that the most discriminating variable at all drill sites is the GR. GR is mainly related to lithology and, in this case, in particular to the clay type and clay content. At all sites, GR has a factor loading of >0.9. Together with the Th and K content, it forms the factor 1 log at all sites. All factor loadings of factor 1 are positive and >0.8; thus the assigned physical or chemical properties show a good positive correlation. Often, the Th/K ratio is taken as an indicator for the clay type (Rider, 1996; Jurado et al., 1997). This means that the factor 1 log is mirroring the lithology with mainly varying clay type and clay content. Because illite has the highest K content among the different clay types (Rider, 1996), borehole sections with high factor 1 values may indicate higher illite concentrations, whereas sections with low factor 1 values may be characteristic of a higher smectite content (in particular within the décollement zones). However, this could also simply mean a lower clay mineral content since factor 1 has a high positive loading for Th as well as K.

At all sites (except Site 1045), the ROMT and TNPH show the highest loadings for factor 2. As expected, the signs of the factor loadings for ROMT and TNPH are opposite: high density sections have low porosity

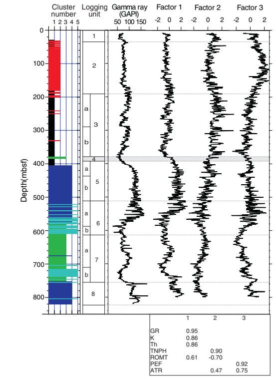
F2. Cluster log, logging units, and downhole logs for Hole 1044A, p. 13.



F3. Cluster log, logging units, and downhole logs for Hole 1045A, p. 14.



F4. Cluster log, logging units, and downhole logs for Hole 1046A, p. 15.



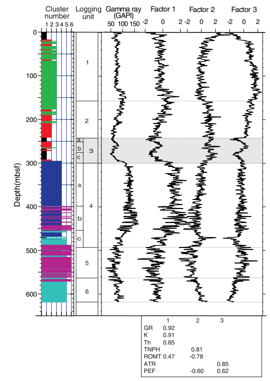
values and vice versa. Thus, factor 2 is mainly responding to the porosity of the formation. After calibration with core sample measurements (e.g., on cores from the nearby Leg 156 sites), the factor 2 log could result in a reliable and true porosity profile. In some cases (Sites 1044, 1047, and 1048), factor 2 is also loaded by PEF. The PEF is closely related to mineralogical composition and, thus, to lithology (Rider, 1996). However, because the PEF is a direct function of atomic number (to the fourth order), pore water and, thus, porosity will also have some influence on PEF in addition to changes in lithology. But PEF is influenced primarily by lithology and only secondarily by porosity. This consequence of basic physics is based on the aggregate atomic number of water, which is much lower than those of rock forming minerals and, when they are mixed together, the combined PEF is controlled by the weight concentration. This means that porosity effects on PEF are much less than seen on either the density or neutron log responses. According to Ellis (1987), kaolinite has a relative low atomic number, but other clay minerals show higher responses to PEF that reflect iron content (having a high atomic number). As ROMT is closely related to PEF, Figure F7 (upper row of crossplots) also mirrors the relation between porosity and PEF. This relation is good above the décollement, but only fair to bad below it.

The deep resistivity ATR, and to a smaller extent PEF, is the main loading for factor 3 at all sites (except Site 1045). Thus, factor 3 is likely related to changes in the electrical properties of the sediments. This might involve several influences such as grain sizes, grain orientation, the presence of conductive minerals, cementation, and varying ionic concentrations within bound water on the clay minerals.

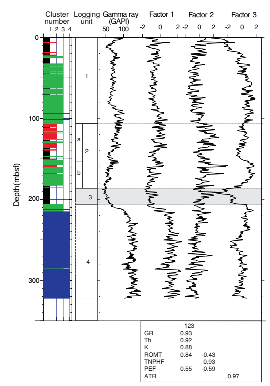
Cluster Logs

Together with the logging units, the cluster logs are shown in color at the left side of Figures F2, F3, F4, F5 and F6. An overview of all cluster logs is given in Figure F7. The mean and standard deviation values for each cluster are given by Moore, Klaus, et al. (1998). Each cluster represents intervals where the physical and chemical rock properties are presumably similar. At all sites, four to six significant clusters could be derived by dendrogram evaluation. In this way, the clusters in the cluster logs can be seen as statistical electrofacies as defined by Serra (1986). This clustering facilitates the subdivision of the borehole into logging units, which can be compared to lithology and porosity. In all cluster logs, the décollement zone is clearly identified by cluster 1 values. As can be derived from Figure F7, cluster 1 is characterized by the lowest density values and by low gamma-ray and PEF values. Clusters 1 and 2 are all above the décollement, whereas clusters 5 and 6 are only below the décollement zone. However, because the cluster logs were calculated individually for each site (because of software constraints), a stratigraphic correlation between the wells by using the clusters has not been possible before now. In the next step, all downhole logging data of Leg 171A will be put together in one data set to perform a consistent suite of cluster logs that can be used to follow up the geological units from well to well. The results compared to the seismic profiles will be further investigated.

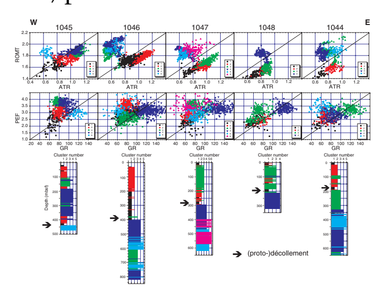
F5. Cluster log, logging units, and downhole logs for Hole 1047A, p. 16.



F6. Cluster log, logging units, and downhole logs for Hole 1048A, p. 17.



F7. Overview of results from multivariate statistical analysis at all sites, p. 18.



RESULTS AND DISCUSSION

By means of the multivariate statistical methods factor and cluster analysis, it was possible to reduce the dimensionality of LWD data from Leg 171A without loss of important information. The resulting set of factor and cluster logs make subsequent evaluations and interpretations much easier. At all sites, the analyses resulted in two or three factor logs that are consistently loaded by comparable factor loadings. We conclude that factor 1 and factor 2 are good proxies for lithology and porosity respectively. Factor 3 possibly contains additional information regarding changes in the electrical properties of the sediments.

No cores were recovered during Leg 171A. However, cores were recovered during Leg 156 and, by design, several of these sites are near Leg 171A sites. The physical properties core measurements from Leg 156 will be used in a future study to calibrate the results from the statistical evaluations of LWD logs from Leg 171A. Obviously, this will enhance the interpretation of the meaning of the factor logs and clusters. Of course, the exact number and choice of input logs for the factor analysis is varying according to the experience of the user and the geology of the logged sequence as well as to the target of the results wanted. But this preconditioning is helpful in improving the factors and making their geological meaning more explicit.

However, the preliminary results from multivariate statistics can be used for first-order interpretation. In Figure F6, all cluster logs are shown together with plots of cross correlations of ATR with ROMT and of GR with PEF. The upper row of crossplots (ATR-ROMT) dramatically shows the change of physical properties behavior above and below the décollement zone. The ATR-ROMT correlation is strong above the décollement zone because changes in porosity greatly exceed changes in lithology. The lithology is essentially homogenous and consists of calcareous clay with minor variations due to clay type, ash content, and structure (fractures, dipping beds, etc.). These small variations in lithology can also be followed up by the lower row of crossplots with the PEF-GR correlations. Following the west-east transect, no big change in the scatter plots can be seen. Therefore, above the décollement zone, the variance in both logs is mainly caused by porosity. Below the décollement zone, there is much greater variability in lithology, which contributes variance to the ATR and ROMT logs in differing manners. For instance, the ATR log is more sensitive to the grain size than is ROMT. Also, there are still effects caused by localized structure. Thus, there is poor correlation between ATR and ROMT below the décollement zone. A similar effect can be seen in the PEF log at Site 1045, for example. The PEF responds primarily to lithology and only secondarily to porosity (Rider, 1996). However, above the décollement zone, PEF shows a strong correlation with ATR. Below the décollement zone, the correlation is very weak. Thus, according to the general compaction trend above the décollement zone, both logs are responding primarily to changes in porosity. Below the décollement zone, PEF responds mainly to lithology variations, whereas ATR still responds primarily to porosity. This is verified by the strong correlation between PEF and ROMT both above and below the décollement zone. In other words, there is clear visual evidence that PEF and ROMT (and GR, Th, and K) are sensitive to lithology, whereas ATR is more sensitive to porosity.

The main objective of using the multivariate methods factor and cluster analysis in this study was twofold:

1. Because there were no cores retrieved during Leg 171A, a reliable lithologic subdivision of the drilled sequences is very difficult to make. Based on the factor and cluster analysis, a classification of the drilled sequences from their physical and chemical properties can be done rapidly and objectively. The factor analysis gives factor logs, which mirror the basic processes behind the physical and chemical properties. By the cluster analysis, similar physical and chemical properties of measured data points are grouped into one cluster, reflecting one lithologic unit.
2. This procedure of objectively grouping measured physical and chemical properties into clusters helped in defining and characterizing logging units. The multivariate statistical methods are helpful tools for reliable, reproducible, and objective definition of logging units, which should be considered for future legs.

ACKNOWLEDGMENTS

This work was supported by the German Science Foundation DFG. Special thanks go to Peter Ireland and Thomas Horton from Anadrill for their excellent work with the LWD tools and also to Captain Ed Oonk and his crew for giving us a calm and efficient cruise. The critical and constructive reviews from John Doveton and Louis Briqueu were very helpful and significantly improved the manuscript.

REFERENCES

Anadrill-Schlumberger, 1993. *Logging While Drilling:* Houston (Schlumberger), document SMP-9160.

Backhaus, K., Erichson, B., Plinke, W., and Weiber, R., 1996. *Multivariate Analysemethoden* (8^{th} ed.): Heidelberg (Springer).

Brown, C.E., 1998. *Applied Multivariate Statistics in Geohydrology and Related Sciences:* Heidelberg (Springer).

Bucheb, J.A., and Evans, H.B., 1994. Some applications of methods used in electrofacies identification. *Log Analyst,* 35:14–26.

Davis, J.C., 1986. *Statistics and Data Analysis in Geology* (2nd ed.): New York (Wiley).

Doveton, J.H., 1994. Geologic log analysis using computer methods. *AAPG Comp. Appl. Geol.,* 2:169.

Elek, I., 1990. Fast porosity estimation by principal component analysis. *Geobyte,* 5:25–34.

Ellis, D.V., 1987. *Well Logging for Earth Scientists:* New York (Elsevier).

Harvey, P.K., Bristow, J.F., and Lovell, M.A., 1990. Mineral transforms and downhole geochemical measurements. *Sci. Drill.,* 1:163–176.

Harvey, P.K., and Lovell, M.A., 1989. Basaltic lithostratigraphy of Ocean Drilling Program Hole 504B. *Nucl. Geophys.,* 3:87–96.

Howarth, R.J., and Sinding-Larsen, R., 1983. Multivariate Analysis. *In* Howarth, R.J. (Ed.), *Statistics and Data Analysis in Geochemical Prospecting. Handbook of Exploration Geology* (Vol. 2): Elsevier (Amsterdam), 207–289.

Jurado, M.J., Moore, J.C., and Goldberg, D., 1997. Comparative logging results in clay-rich lithologies on the Barbados Ridge. *In* Shipley, T.H., Ogawa, Y., Blum, P., and Bahr, J.M. (Eds.), *Proc. ODP, Sci. Results,* 156: College Station, TX (Ocean Drilling Program), 321–334.

Kovach, W.L., 1998. *MVSP – A Multivariate Statistical Package for Windows, ver. 3.0.* Kovach Computing Services, Pentraeth, Wales, U.K.

Moore, J.C., Klaus, A., Bangs, N.L., Bekins, B., Bücker, C.J., Brückmann, W., Erickson, N.E., Horton, T., Ireland, P., Major, C.O., Peacock, S., Saito, S., Screaton, E.J., Shimeld, J.W., Stauffer, P.H., Taymaz, T., Teas, P.A., and Tokunaga, T., 1998. Consolidation patterns during initiation and evolution of a plate-boundary décollement zone: Northern Barbados accretionary prism. *Geology,* 26:811–814.

Moore, J.C., Klaus, A., et al., 1998. *Proc. ODP, Init. Repts.,* 171A: College Station, TX (Ocean Drilling Program).

Rider, M., 1996. *The Geological Interpretation of Well Logs* (2nd ed.): Caithness (Whittles Publishing).

Serra, O., 1986. *Fundamentals of Well-Log Interpretation* (Vol. 2): *The Interpretation of Logging Data*. Dev. Pet. Sci., 15B.

Shipboard Scientific Party, 1998. Explanatory notes. *In* Moore, J.C., Klaus, A., et al., *Proc. ODP, Init. Repts.,* 171A: College Station, TX (Ocean Drilling Program), 11–15.

Shipley, T.H., Moore, G.F., Bangs, N.L., Moore, J.C., and Stoffa, P.L., 1994. Seismically inferred dilatancy distribution, northern Barbados Ridge décollement: implications for fluid migration and fault strength. *Geology,* 22:411–414.

Tabachnick, B.G., and Fidell, L.S., 1989. *Using Multivariate Statistics:* San Francisco (Harper and Row).

Theys, P.P., 1991. *Log Data Acquisition and Quality Control:* Paris (Ed. Technip).

Figure F1. Seismic depth section extending from west of Site 1045 to east of Site 1044. The depth profiles of density in the seismic section are reflecting the décollement zone by low density values. The décollement zone and the depths of drilling are indicated.

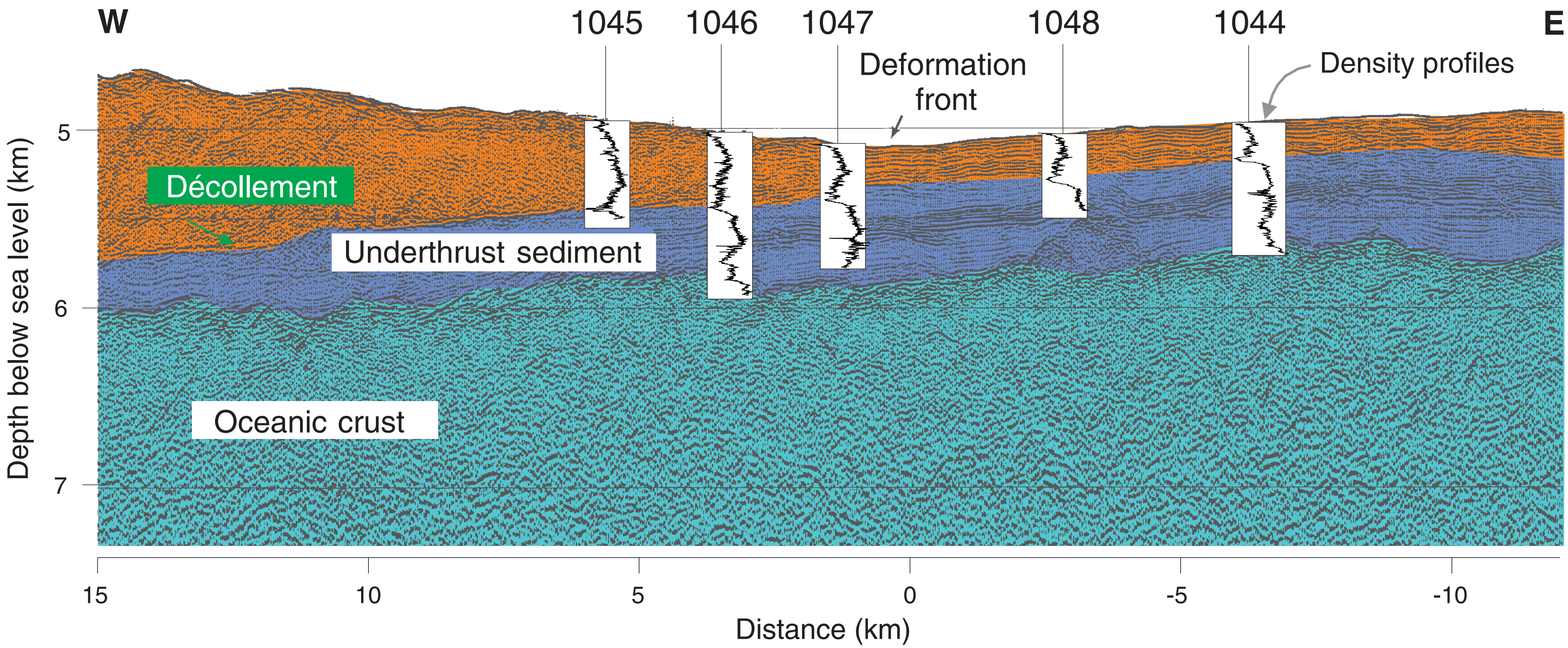


Figure F2. Cluster log, logging units, and downhole logs for Hole 1044A. Factor 1, factor 2, and factor 3 are the factor logs as derived by factor analysis. Based on these three factor logs, the cluster log was calculated. The downhole logging units were defined by the help of the cluster log. In the bottom part of the diagram, the factor loadings with values >0.4 are shown. The factor 1 log is mainly related to lithology (changing clay content and/or clay type, high loading with GR, Th, and K), whereas the factor 2 log is related to porosity (high loading of TNPH and ROMT). GR = gamma ray; K = potassium; Th = thorium; U = uranium; ROMT = density; TNPH = neutron porosity; ATR = electrical resistivity; PEF = photoelectric effect.

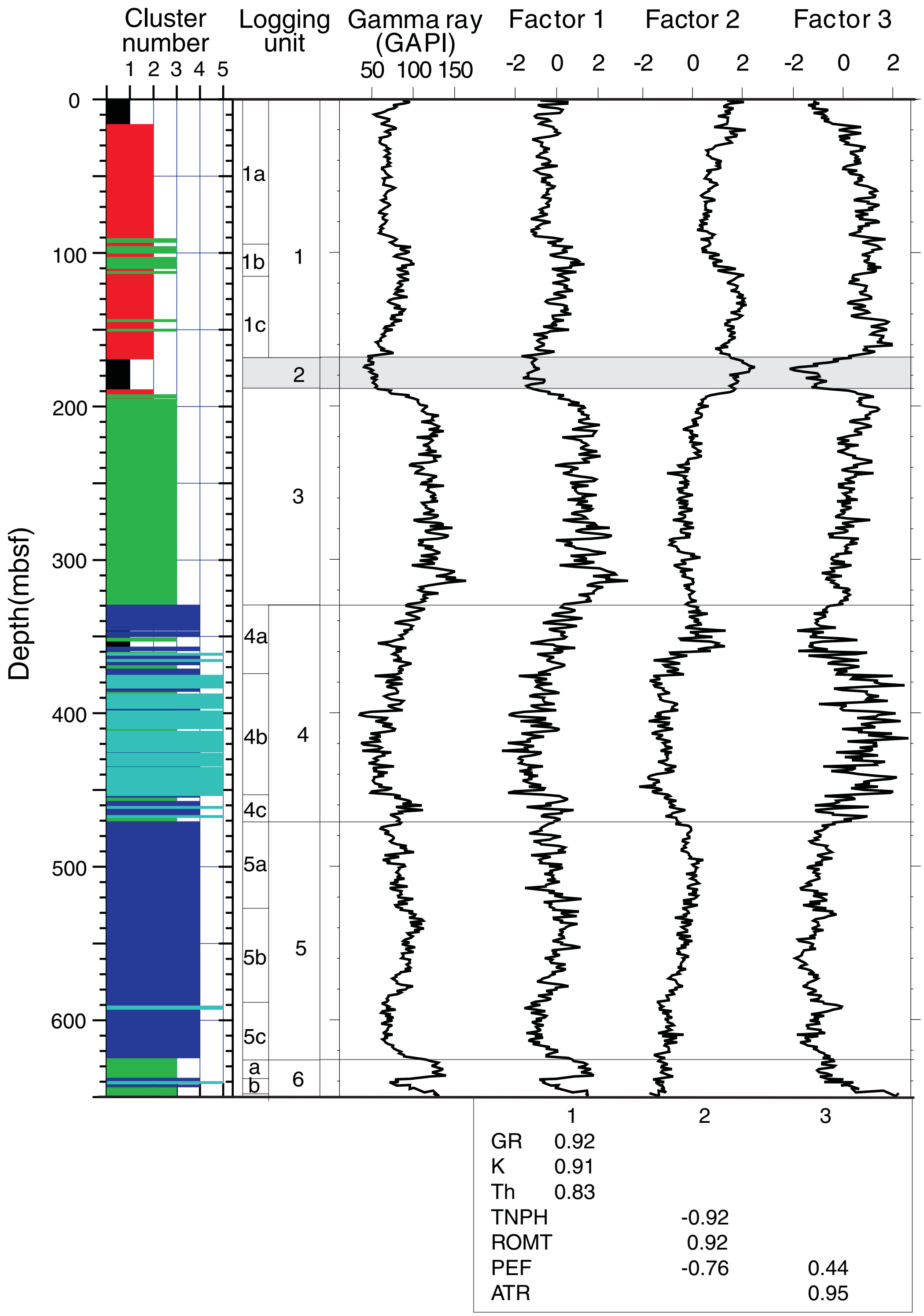


Figure F3. Cluster log, logging units, and downhole logs for Hole 1045A. Factor 1 and factor 2 are the factor logs as derived by factor analysis. Based on these two factor logs, the cluster log was calculated. The downhole logging units were defined by the help of the cluster log. In the bottom part of the diagram, the factor loadings with values >0.4 are shown. The factor 1 log is mainly related to lithology (changing clay content and/or clay type), whereas the factor 2 log is related to porosity (high loading of TNPH and ROMT).

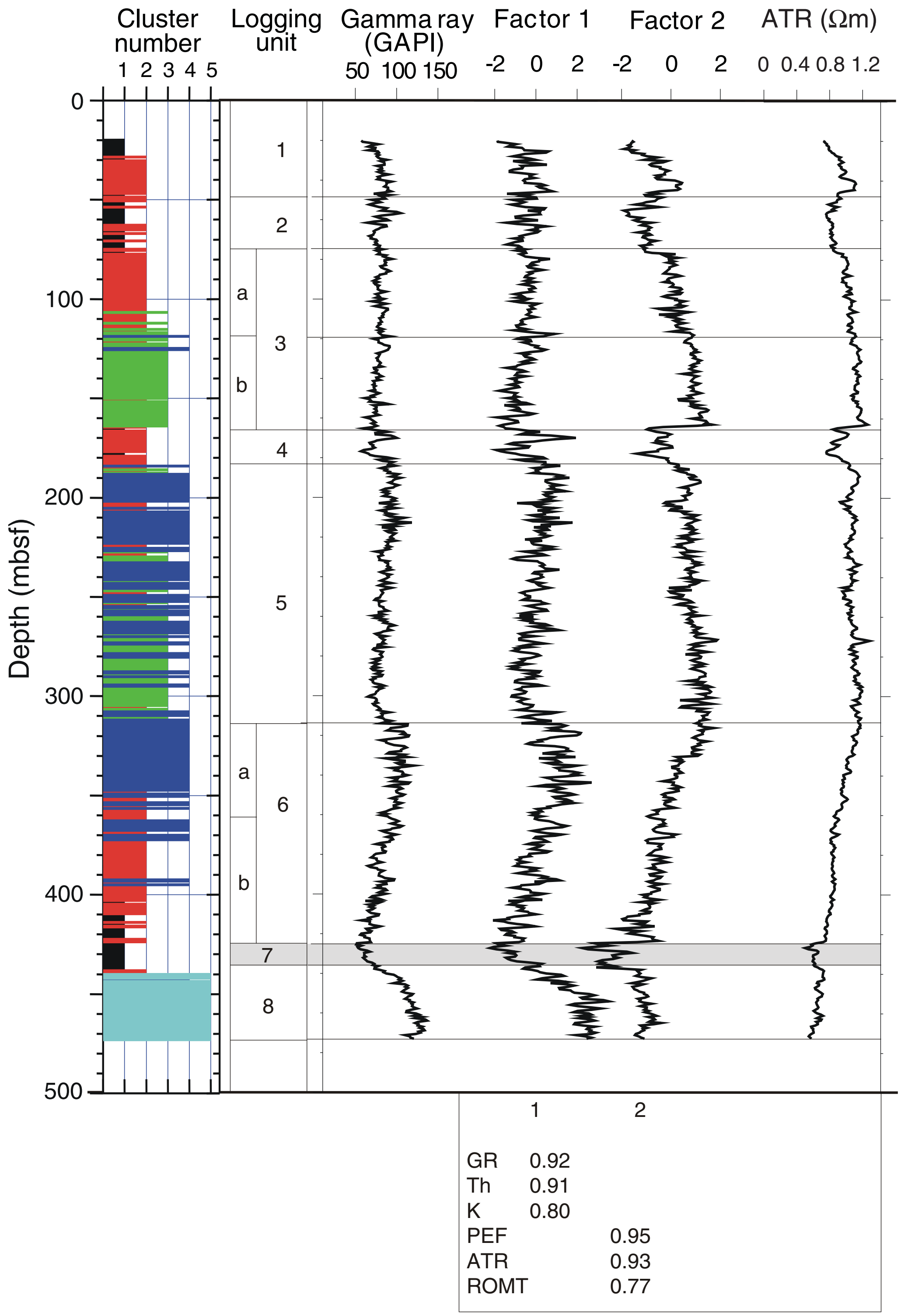
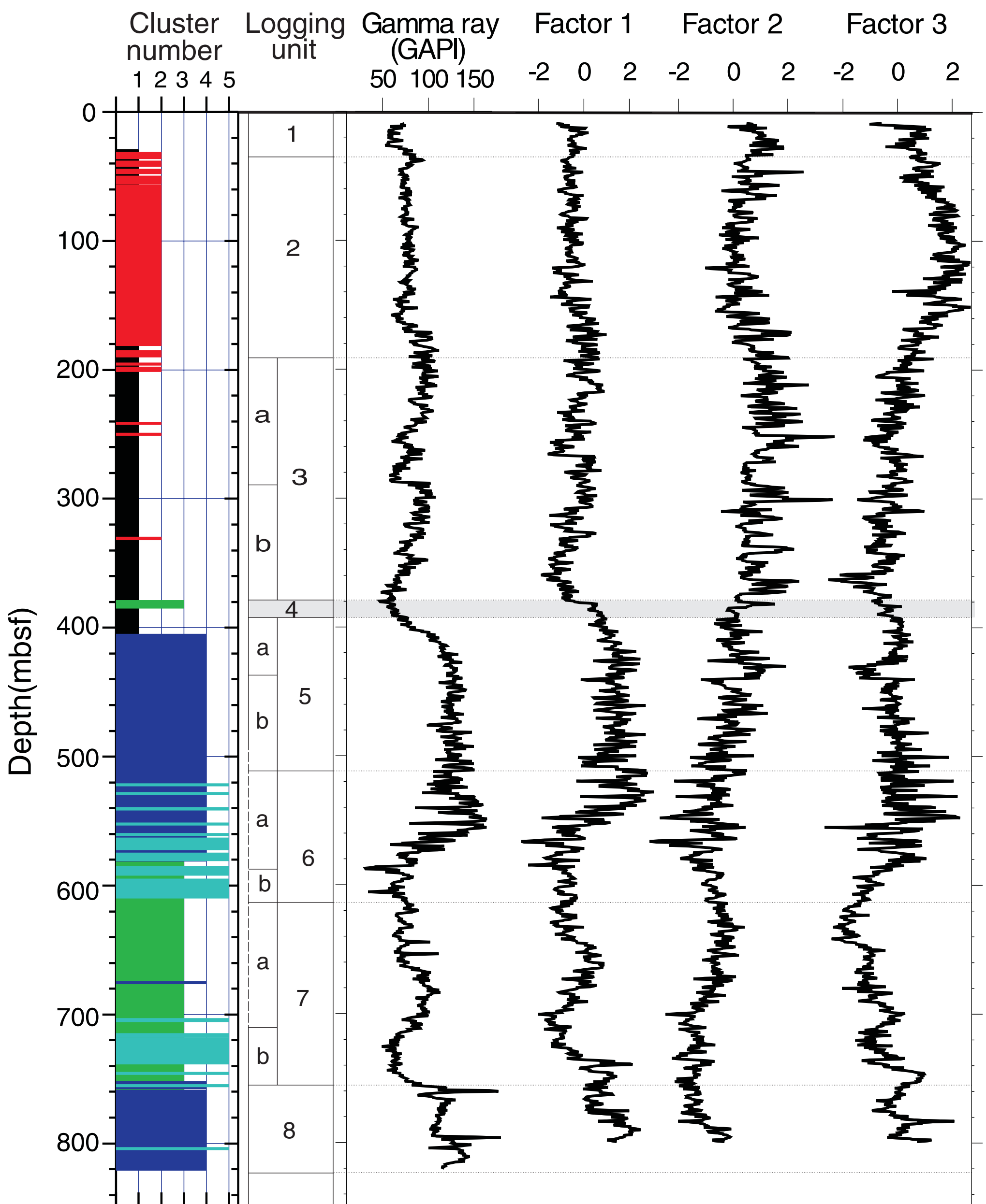


Figure F4. Cluster log, logging units, and downhole logs for Hole 1046A. Factor 1, factor 2, and factor 3 are the factor logs as derived by factor analysis. Based on these three factor logs, the cluster log was calculated. The downhole logging units were defined by the help of the cluster log. In the bottom part of the diagram, the factor loadings with values >0.4 are shown. The factor 1 log is mainly related to lithology (changing clay content and/or clay type), whereas the factor 2 log is related to porosity (high loading of TNPH and ROMT).



	1	2	3
GR	0.95		
K	0.86		
Th	0.86		
TNPH		0.90	
ROMT	0.61	-0.70	
PEF			0.92
ATR		0.47	0.75

Figure F5. Cluster log, logging units, and downhole logs for Hole 1047A. Factor 1 and factor 2 are the factor logs as derived by factor analysis. Based on these three factor logs, the cluster log was calculated. The downhole logging units were defined by the help of the cluster log. In the bottom part of the diagram, the factor loadings with values >0.4 are shown. The factor 1 log is mainly related to lithology (changing clay content and/or clay type), whereas the factor 2 log is related to porosity (high loading of TNPH and ROMT).

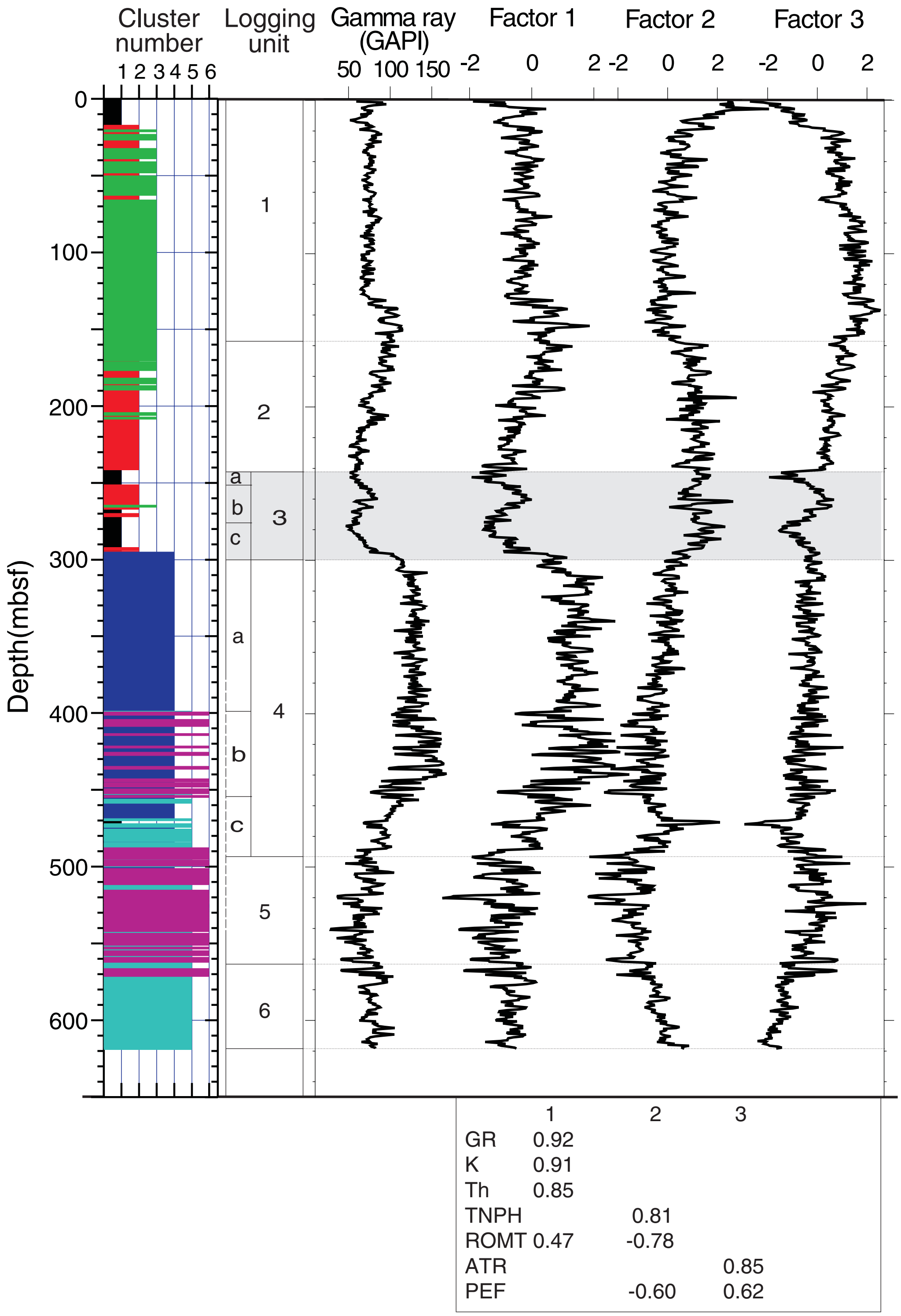
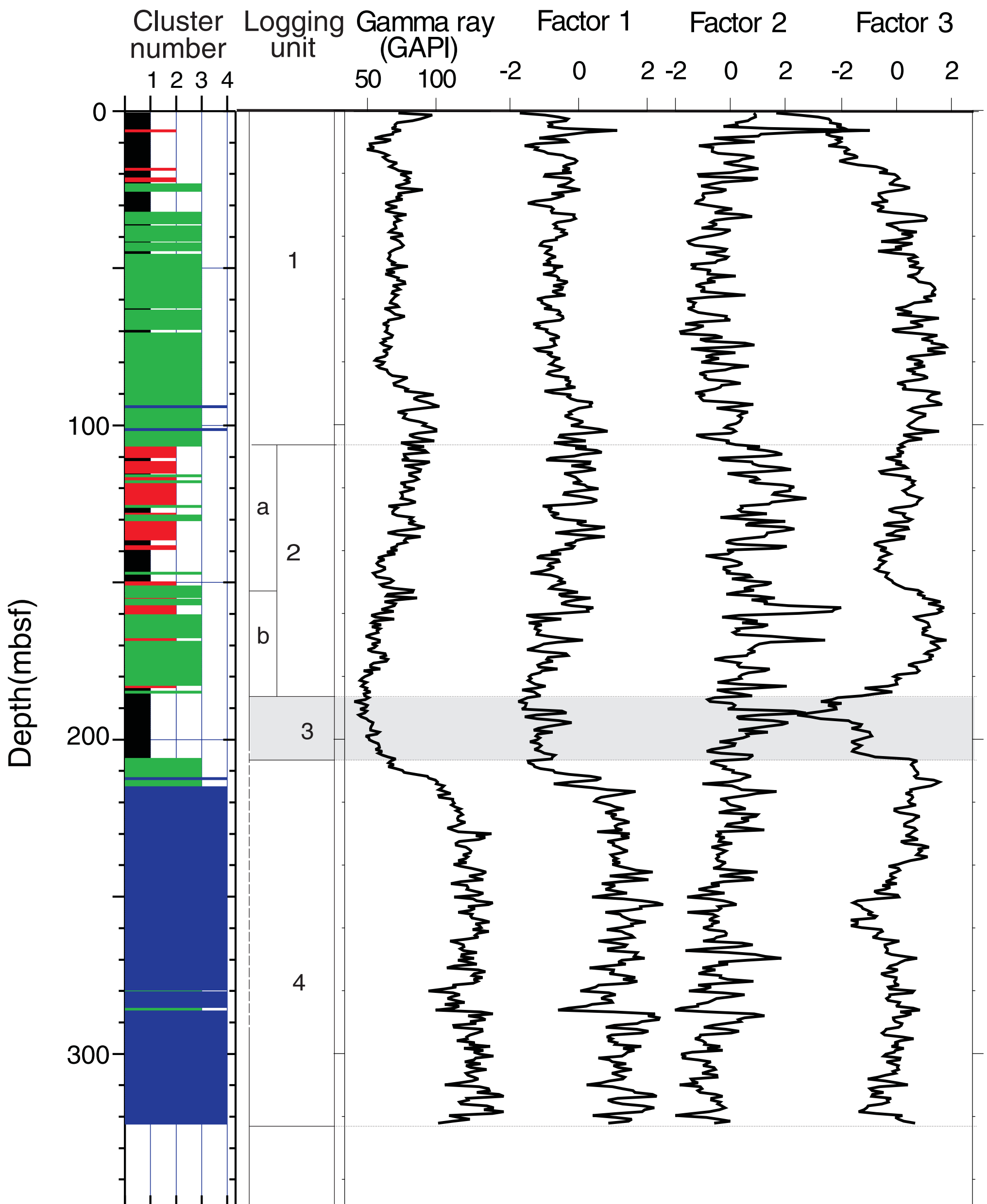


Figure F6. Cluster log, logging units, and downhole logs for Hole 1048A. Factor 1 and factor 2 are the factor logs as derived by factor analysis. Based on these factor logs, the cluster log was calculated. The downhole logging units were defined by the help of the cluster log. In the bottom part of the diagram, the factor loadings with values >0.4 are shown. The factor 1 log is mainly related to lithology (changing clay content and/or clay type), whereas the factor 2 log is related to porosity (high loading of TNPH and ROMT).



	1	2	3
GR	0.93		
Th	0.92		
K	0.88		
ROMT	0.84	-0.43	
TNPHF		0.93	
PEF	0.55	-0.59	
ATR			0.97

Figure F7. Comparative overview of results from multivariate statistical analysis at all sites from Leg 171A along the west-east transect. The cluster logs in the lower part of the figure are the same as in Figures F2, p. 13, F3, p. 14, F4, p. 15, F5, p. 16, and F6, p. 17, and are shown for orientation reasons. The (proto-)décollement zone is indicated by an arrow. The colors in the crossplots in the upper part of the figure correspond to the colors in the cluster logs. By using the colors, the depths trends of the statistical electrofacies can be followed in the crossplots. The upper row of crossplots shows the density (ROMT)-resistivity (ATR) relation. At all sites, a linear ROMT-ATR relation can be seen for the depth section above the décollement. This linear relation is lost below the décollement, dramatically showing the change of physical properties above and below the décollement. The diagonal lines in the scatter plots are shown as reference lines and can be used for comparison purposes. The lower row of crossplots shows the photoelectric effect (PEF)-gamma ray (GR) relation. PEF and GR are mainly related to lithology. In all crossplots the dots are plotting in more or less the same region. Thus, the lithology does not change much with depth or distance from the deformation front. For further explanation see Results and Discussion, p. 9. (Figure shown on next page.)

Figure F7 (continued). (Caption on previous page.)

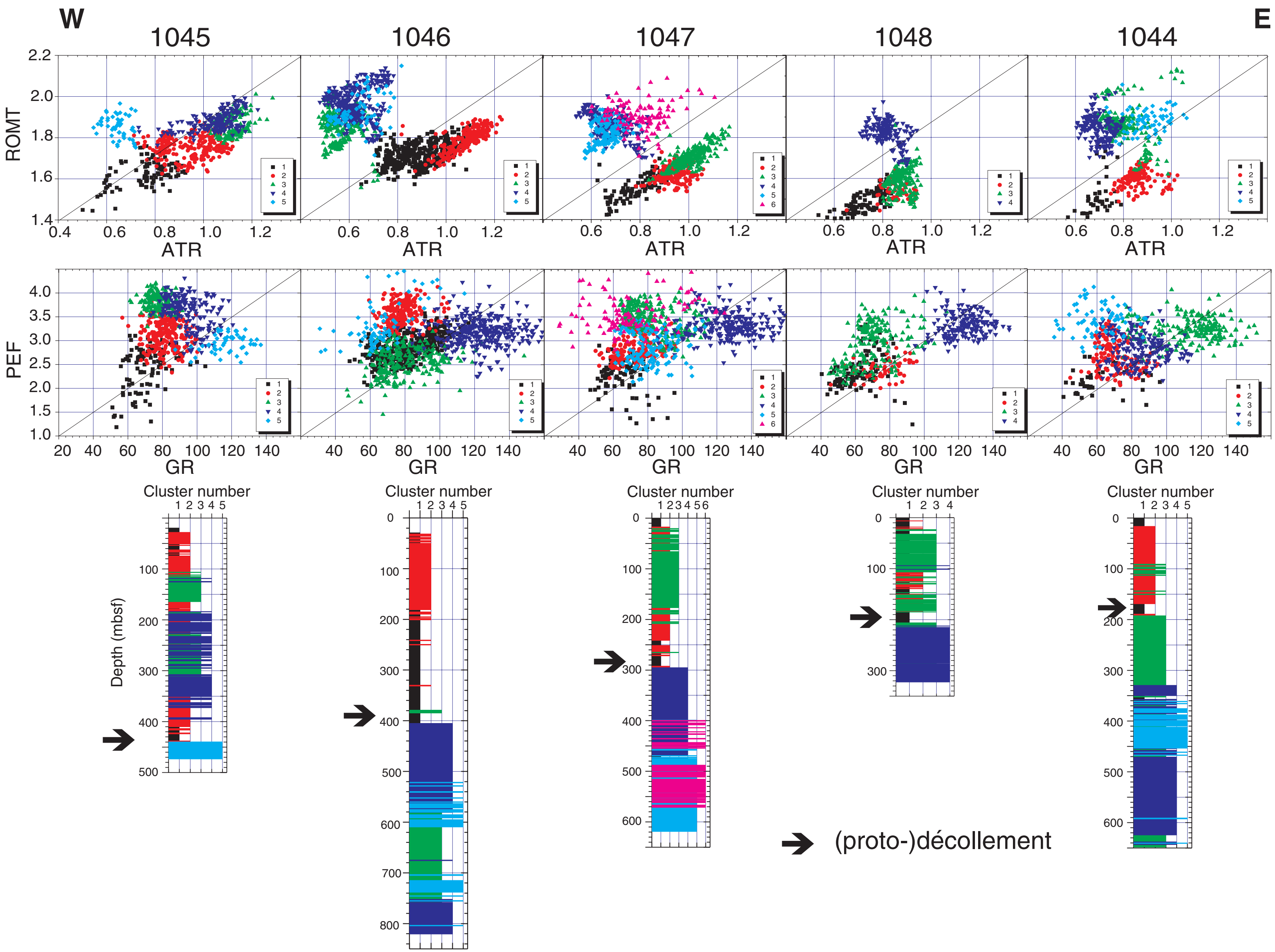


Table T1. Output and results of factor analysis for Site 1044.

Factor	Eigenvalue	Variance percent	Percent cumulative
*1	3.14	44.9	44.9
*2	1.84	26.3	71.2
*3	1.22	17.4	88.6
4	0.35	5.1	93.6
5	0.22	3.1	96.7
6	0.15	2.1	98.9
7	0.07	1.1	100.0

Notes: The number of valid cases (643) reflects the number of data in the borehole log. For each factor the eigenvalues and the amount of explained variances are given. Bold * = eigenvalues that are assumed important because they are ≥1. Figures **F2,** p. 13, **F3,** p. 14, **F4,** p. 15, **F5,** p. 16, and **F6,** p. 17, also have factors indicated by an * and are included in the following cluster analysis.

Table T2. Site 1044 Varimax factor loadings and communalities.

	Factors			
	1	2	3	Communality
GR	0.92	–0.19	–0.08	0.89
K	0.91	0.16	0.15	0.87
Th	0.83	–0.36	–0.21	0.87
TNPH	0.12	0.92	0.23	0.91
ROMT	0.30	–0.92	–0.06	0.94
PEF	0.19	–0.76	0.44	0.82
ATR	–0.08	0.05	0.95	0.91
Squaresum:	2.51	2.45	1.22	6.20
Variance (%):	35.9	35.1	17.5	88.6

Notes: This table gives the corresponding factor landings, the communality, and the total amount of explained variance. The factor loadings are simply the weights loaded on the factors. Factor loadings >0.4 are shown in bold. The sum of the factor loadings squared (Squaresum) is equal to the eigenvalue, which is the variance explained by a factor. Three factors have an eigenvalue >1; the total explained variance is 88.6%.

Table T3. Output and results of factor analysis for Site 1045.

Factor	Eigenvalue	Variance (%)	Cumulative (%)
***1**	2.8	47.3	47.3
***2**	2.10	35.1	82.3
3	0.49	8.2	90.6
4	0.25	4.2	94.9
5	0.19	3.1	98.0
6	0.11	1.9	100.0

Notes: Number of valid cases: 744. Two factors have an eigenvalue >1; the total explained variance is 82.3%. See Table T1, p. 20, for explanation.

Table T4. Site 1045 Varimax factor loadings and communalities.

	Factors		
	1	2	Communality
GR	0.91	0.06	0.84
Th	0.90	–0.09	0.83
K	0.80	0.18	0.67
PEF	0.06	0.95	0.91
ATR	–0.17	0.92	0.88
ROMT	0.44	0.77	0.79
Squaresum:	2.53	2.40	4.94
Variance (%):	42.2	40.1	82.3

Notes: The total explained variance is 82.3%. See Table T2, p. 21, for explanation.

Table T5. Output and results of factor analysis for Site 1046.

Factor	Eigenvalue	Variance (%)	Cumulative (%)
***1**	3.23	46.2	46.2
***2**	1.66	23.7	69.9
***3**	1.21	17.3	87.3
4	0.43	6.1	93.4
5	0.20	2.9	96.4
6	0.13	1.9	98.3
7	0.11	1.6	100.0

Notes: Three factors have an eigenvalue > 1; the total explained variance is 87.3%. See Table T1, p. 20, for an explanation.

Table T6. Site 1046 Varimax factor loadings and communalities.

	Factors			Commuality
	1	2	3	
GR	0.95	–0.10	0.00	0.91
K	0.86	0.22	0.18	0.82
Th	0.86	–0.30	–0.19	0.87
TNPH	0.09	0.89	0.02	0.81
ROMT	0.60	–0.70	0.07	0.87
PEF	0.20	–0.19	0.92	0.92
ATR	–0.30	0.47	0.75	0.88
Squaresum:	2.90	1.71	1.49	6.11
Variance (%):	41.4	24.5	21.2	87.3

Notes: The total explained variance is 87.3%. See Table T2, p. 21, for explanation.

Table T7. Output and results of factor analysis for Site 1047.

Factor	Eigenvalue	Variance (%)	Cumulative (%)
***1**	3.34	47.7	47.7
***2**	1.44	20.6	68.3
***3**	1.13	16.1	84.5
4	0.61	8.8	93.3
5	0.18	2.6	96.0
6	0.15	2.1	98.1
7	0.12	1.8	100.0

Notes: Number of valid cases: 1014. Three factors have an eigenvalue >1; the total explained variance is 84.5%. See Table T1, p. 20, for explanation.

Table T8. Site 1047 Varimax factor loadings and communalities.

	Factors			
	1	2	3	Communality
GR	0.92	–0.14	–0.16	0.90
K	0.90	0.07	0.11	0.84
Th	0.85	–0.20	–0.31	0.87
TNPH	0.17	0.81	0.09	0.69
ROMT	0.46	–0.78	–0.08	0.83
ATR	–0.33	0.22	0.85	0.88
PEF	0.35	–0.59	0.62	0.87
Squaresum:	2.89	1.74	1.27	5.91
Variance (%):	41.3	24.9	18.1	84.5

Notes: The total explained variance is 84.5%. See Table T2, p. 21, for explanation.

Table T9. Output and results of factor analysis for Site 1048.

Factor	Eigenvalue	Variance (%)	Cumulative (%)
***1**	4.38	62.7	62.7
***2**	1.10	15.8	78.5
***3**	0.73	10.4	89.0
4	0.37	5.3	94.3
5	0.24	3.5	97.9
6	0.11	1.5	99.5
7	0.03	0.4	100.0

Notes: The number of valid cases is 528. Three factors have an eigenvalue >1; the total explained variance is 89.0%. See Table **T1,** p. 20, for explanation.

Table T10. Site 1048 Varimax factor loadings and communalities.

	Factors			
	1	2	3	Communality
GR	0.93	–0.25	0.00	0.93
Th	0.91	–0.20	–0.02	0.88
K	0.87	–0.16	0.00	0.79
ROMT	0.83	–0.43	0.20	0.93
TNPH	–0.24	0.92	0.08	0.92
PEF	0.55	–0.58	0.39	0.80
ATR	0.00	0.00	0.97	0.94
Squaresum:	3.55	1.52	1.15	6.23
Variance (%):	50.7	21.8	16.4	89.0

Notes: The total explained variance is 89.0%. See Table T2, p. 21, for explanation.